

# Exploratory pathway analysis

Citation for published version (APA):

Kelder, T. A. J. (2011). *Exploratory pathway analysis*. [Doctoral Thesis, Maastricht University]. Maastricht University. <https://doi.org/10.26481/dis.20110708tk>

**Document status and date:**

Published: 01/01/2011

**DOI:**

[10.26481/dis.20110708tk](https://doi.org/10.26481/dis.20110708tk)

**Document Version:**

Publisher's PDF, also known as Version of record

**Please check the document version of this publication:**

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

**Take down policy**

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

# Samenvatting

Een belangrijk doel van biologisch onderzoek is om beter te begrijpen hoe interacties tussen verschillende moleculen het functioneren en disfunctioneren van organismen bepalen. Deze mechanismen zijn vaak zeer complex. Ter illustratie: Het humane DNA bestaat uit ongeveer drie miljard base paren die coderen voor meer dan negenentwintig duizend eiwitten. Deze eiwitten komen voor in verschillende hoeveelheden en varianten, die afhankelijk zijn van omgevingsfactoren, cellulaire locatie en onderlinge interacties.

Er is al veel bekend over deze interacties en mechanismen in de context van verschillende organismen en ziekten. Een veelgebruikte manier om deze kennis weer te geven is het concept biologisch pathway. Een typisch pathway beschrijft een set moleculaire entiteiten (e.g. genen, eiwitten of metabolieten), hun onderlinge relaties of interacties, en variaties in tijd of locatie. De verschillende stappen in een pathway worden vaak op een grafische manier weergegeven in de vorm van een diagram. In een pathway wordt onze kennis over de moleculaire biologie op een functioneel niveau beschreven en georganiseerd. Over de jaren is er een canon van pathways ontstaan die de werking van veel voorkomende biologische mechanismen samenvatten.

Dankzij de ontwikkeling van nieuwe technieken is het de laatste decennia mogelijk geworden om relatief gemakkelijk grote hoeveelheden experimentele data te genereren. De zogenoemde 'high-throughput' of 'omics' technieken stellen biologen in staat om metingen te doen op duizenden entiteiten binnen een enkel experiment. Hierdoor kan bijvoorbeeld de expressie van alle genen of de aanwezigheid van vele eiwitten tegelijk gemeten worden. Door deze metingen uit te voeren op verschillende tijdstippen, of door het vergelijken van verschillende cel typen, kan bepaald worden welke veranderingen er optreden op moleculair niveau. Door het grote bereik maar relatief hoge ruis in dit type experimenten zijn deze uitermate geschikt voor exploratieve analyse. Dit soort analyse heeft als doel om gerichte vragen en hypothesen te genereren die vervolgens getest kunnen worden in meer gerichte experimenten.

De hoeveelheden experimentele data zijn zo groot dat ze niet meer efficiënt met de hand geanalyseerd kunnen worden. Daarom worden er binnen de bioinformatica algoritmen en computer software ontworpen en toegepast om het interpreteren van deze data mogelijk te maken. Een belangrijke stap in een analyse is het combineren van de data met de kennis die we al hebben, om zo tot nieuwe biologische inzichten te komen. Dit proefschrift focust op dat deel van de analyse, waarbij pathways worden gebruikt om bestaande kennis te integreren met bioinformatica toepassingen.

Het doel van dit proefschrift is om exploratieve data analyse te verbeteren door het gebruiken van biologische pathways. Hoofdstuk 2 bevat een korte samenvatting van bestaande pathway analyse technieken. Twee veelgebruikte toepassingen zijn het visualiseren van data op pathway diagrammen en het uitvoeren van over-representatie analyse, waarbij wordt gezocht naar pathways die waarschijnlijk betrokken zijn in het bestudeerde biologische mechanisme. Verder zijn een aantal kenmerken en vereisten van exploratieve data analyse beschreven, die verder in het proefschrift worden gebruikt om analyse methoden te verbeteren en toe te passen.

Een van de vereisten om pathways te kunnen gebruiken in data analyse is dat onze huidige kennis over moleculaire biologie wordt gerepresenteerd als pathways. Het verzamelen van biologische pathways is een immense taak, omdat deze niet automatisch kunnen worden gegenereerd,

maar handmatig moeten worden gecompileerd uit kennis verspreid over wetenschappelijke literatuur en verschillende domein experts. De pathway database WikiPathways die is beschreven in hoofdstuk 3 is een unieke benadering om deze grote hoeveelheden biologische kennis te verzamelen in een vorm die gebruikt kan worden in bioinformatica toepassingen. WikiPathways is een wiki waarbij, zoals op Wikipedia, alle informatie direct te bewerken is door gebruikers. Dit geeft de wetenschappelijke gemeenschap zelf controle om de informatie te verbeteren in plaats van dit aan een kleine groep curators over te laten. Hiermee biedt WikiPathways op termijn een beter schaalbare oplossing om de exponentiële groei van nieuwe kennis aan te kunnen. Een vereiste hiervoor is wel dat het aantal actieve gebruikers groot genoeg is om een goede kwaliteit en actuele inhoud te verkrijgen. Twee voorbeelden van punten die dit moeten bevorderen zijn een directe bruikbaarheid van de pathway informatie in onderzoek en een laagdrempelige interface om pathway informatie in te voeren in de vorm van een intuïtief werkend tekenprogramma. Gezien het kort bestaan van WikiPathways is het moeilijk te zeggen of een gebruikersgroep van voldoende grootte bereikt kan worden, echter zowel de inhoud als het aantal actieve gebruikers laten een positieve groei zien.

In hoofdstuk 4 wordt een webservice beschreven die het mogelijk maakt effectiever gebruik te maken van pathway informatie in exploratieve data analyse. Deze webservice geeft computerprogramma's direct toegang tot de informatie op WikiPathways. Dit heeft diverse toepassingen. Ten eerste wordt het hierdoor mogelijk om pathways te integreren in verschillende bioinformatica software. Zo is het bijvoorbeeld mogelijk om vanuit het netwerk visualisatie programma Cytoscape direct de nieuwste versie van een pathway op WikiPathways te in laden. Ook geeft het de mogelijkheid om bijvoorbeeld op de NCBI website informatie over een specifiek gen uit te breiden met de verschillende pathways waarin het een rol speelt. Ten tweede maakt de webservice het ook makkelijker de pathways direct te integreren in een analyse script of workflow en eventueel te combineren met andere typen informatie of publieke datasets die ook via webservices beschikbaar zijn. Als laatste maakt de webservice het ook mogelijk om curatie taken op WikiPathways te assisteren. Zo draaien er bijvoorbeeld voortdurend computerprogramma's op de achtergrond die bijdragen van gebruikers controleren op veel voorkomende fouten en deze terugrapporteren zodat ze verbeterd kunnen worden.

Hoofdstuk 5 geeft een voorbeeld van een exploratieve data analyse die is opgebouwd rondom pathway informatie. Het doel van deze analyse is om inzicht te krijgen in mogelijke interacties tussen pathways in de context van insuline resistentie. Zoals veel complexe ziekten spelen bij insuline resistentie waarschijnlijk verschillende pathways een rol en is het dus nuttig om te kijken naar het samenspel tussen pathways. Veel bestaande pathway gebaseerde methoden gebruiken pathways als geïsoleerde entiteiten, dus biedt deze analyse een nieuw perspectief op de data. De analyse gaat uit van een gen expressie dataset gemeten in lever weefsel van gezonde en insuline resistente muizen, voorafgaand aan en op verschillende tijdstippen na het uitvoeren van een glucose tolerantie test. Op basis van een netwerk met interacties tussen genen en hun eiwit producten wordt tussen elk paar pathways een set kortste paden gezocht waarbij de lengte van een pad de som is van de gewichten van de verbindende eiwit interacties. Interacties tussen eiwitten waarvan de coderende genen differentieel tot expressie komen krijgen een lager gewicht toegewezen, waardoor paden met veel veranderde genen een kortere lengte krijgen. Op basis van de aanname dat hoe meer en kortere paden tussen twee pathways gevonden worden, hoe waarschijnlijker het is dat deze pathways een interactie

aangaan, kan dan een netwerk van pathway interacties worden afgeleid. Op deze manier zijn diverse mogelijke pathway interacties gevonden voor verschillende vergelijkingen van de gen expressie tussen gezonde en insuline resistente groepen en tussen de verschillende tijdstippen. Bovendien is er ingezoomd op de onderliggende eiwit interacties om mogelijke mechanismen en eiwitten die betrokken zijn bij communicatie tussen pathways te bestuderen. Dit heeft tot nieuwe inzichten en hypothesen geleid over het mogelijk samenspel tussen verschillende pathways in insuline resistentie en hoe deze verandert als gevolg van een glucose stimulans.

Tezamen heeft het onderzoek zoals beschreven in dit proefschrift op een aantal punten bijgedragen aan de verbetering van exploratieve data analyse met behulp van biologische pathways. Er is een platform gecreëerd voor het verzamelen en verbeteren van pathway informatie die geschikt is voor gebruik in bioinformatica toepassingen. Er is een infrastructuur gebouwd om deze informatie programmatisch toegankelijk te maken en te integreren in analyses. Tot slot is gedemonstreerd hoe de ontwikkelde hulpmiddelen rondom pathway informatie gebruikt kunnen worden om originele analyse methoden te ontwikkelen die nieuwe perspectieven bieden op een bestaande dataset.